RADIATION RESEARCH **166**, 303–312 (2006)
0033-7587/06 $15.00

Some Statistical Implications of Dose Uncertainty in Radiation Dose–Response Analyses

Daniel W. Schafer[a,1] and Ethel S. Gilbert[b]

[a] *Department of Statistics, Oregon State University, Corvallis, Oregon; and* [b] *Radiation Epidemiology Branch, National Cancer Institute, Bethesda, Maryland*

Schafer, D. W. and Gilbert, E. S. Some Statistical Implications of Dose Uncertainty in Radiation Dose–Response Analyses. *Radiat. Res.* 166, 303–312 (2006).

Statistical dose–response analyses in radiation epidemiology can produce misleading results if they fail to account for radiation dose uncertainties. While dosimetries may differ substantially depending on the ways in which the subjects were exposed, the statistical problems typically involve a predominantly linear dose–response curve, multiple sources of uncertainty, and uncertainty magnitudes that are best characterized as proportional rather than additive. We discuss some basic statistical issues in this setting, including the bias and shape distortion induced by classical and Berkson uncertainties, the effect of uncertain dose-prediction model parameters on estimated dose–response curves, and some notes on statistical methods for dose–response estimation in the presence of radiation dose uncertainties.

INTRODUCTION

Radiation epidemiology is set apart by extensive efforts to estimate a biologically relevant quantity called ''dose'' and to use resulting dose estimates to examine and quantify dose–response relationships. As is evident in other papers in this volume, dose estimates are subject to many sources of uncertainty. Results from statistical analyses that incorrectly assume doses are measured precisely may produce misleading results, including bias in estimated regression coefficients, distortion of the shape of the dose response, underestimation of uncertainty in parameter estimates, and bias in the estimates of effect modification (*1–6*). In addition, statistical power may be overestimated in the planning stages of a study (*7*).

This paper focuses on statistical modeling of dose uncertainties, the bias and shape distortion in the estimation of linear and linear-quadratic dose–response models due to ''multiplicative'' dose uncertainties, including uncertainties that are common to many subjects, and some comments on statistical methods that account for dose uncertainties. Several additional aspects of dose uncertainty that are important but that will not be discussed here are the following: (1) Tests and confidence intervals may be affected by dose uncertainty, even when no significant dose–response bias is induced. (2) If the magnitude and nature of uncertainties in dose differ among subgroups or studies, their distorting effects may also differ, leading to biased comparisons of the dose response among the different subgroups or studies. (3) Uncertainties in one variable (such as the amount that a subject has smoked in his lifetime) can induce biases in the estimated coefficients of another variable with which it is correlated (which could be dose of radiation).

[1] Address for correspondence: Department of Statistics, 44 Kidder Hall, Oregon State University, Corvallis, OR 97331-4606; e-mail: schafer@science.oregonstate.edu.

MODELS RELATING OBSERVED AND TRUE DOSE

A typical dose–response model in radiation epidemiology specifies the relative risk of some cancer to depend on a linear or possibly linear-quadratic function of organ-specific dose of radiation, possibly with modifying effects of sex, age at exposure, and attained age. If the doses were known precisely, then common statistical tools for binary responses or censored survival times could be used to make inferences about the models.

We use the term *observed dose* to represent the available estimate of *true dose*. As a starting point, we write

$$\textit{observed dose} = \textit{true dose} + \textit{error}, \qquad (1)$$

where the term ''*error*'', which is commonly used for such deviations in statistics, simply denotes the (unknowable) difference between the *observed* and *true dose*. For the types of problems of interest here—in which the typical magnitude of the *error* is a percentage of the *true dose*—a multiplicative version is usually more convenient, *observed dose* = *true dose* × *error factor*, where *error factor* is a ratio rather than a difference. This may be expressed as Eq. (1) on a log scale,

$$\log(\textit{observed dose}) = \log(\textit{true dose}) + \log(\textit{error factor}), \qquad (2)$$

so many of the basic issues can be addressed in terms of Eq. (1).

The presence of error-prone *observed doses* obviously complicates inference about the dose–response relationship. A key requirement for ascertaining the potential consequences and appropriate remedies is an understanding, through a probability model, of the way in which *true dose* and *observed dose* are related. Unfortunately, there are several realistic possibilities with different consequences and remedies and no data analysis techniques for deciding which is appropriate. This leaves careful consideration of the dosimetry process and a good understanding of the possible probability models as the only tools for model development. A primary distinction in this regard is between classical and Berkson models for uncertainty.

Classical and Berkson Models for a Single Source of Dose Uncertainty

Two versions of the additive error model in Eq. (1) are the following:

Classical measurement error model:

$$\textit{observed dose} = \textit{true dose} + \textit{measurement error}, \quad (3)$$

in which *measurement error* is a random variable with mean 0, independent of *true dose*, and

Berkson error model:

$$\textit{true dose} = \textit{observed dose} + \textit{individual peculiarity}, \quad (4)$$

in which *individual peculiarity* is a random variable with mean 0, independent of *observed dose*. Notice that the distinction is in whether the discrepancy between the *true* and *observed doses* is independent of *true dose* or *observed dose*. It cannot be independent of both. Since statistical independence plays a crucial role, its definition is reviewed in Appendix A.

The classical measurement error model is ''classical'' in its appropriateness for a true *measurement error* stemming from an imprecise measuring instrument, such as an imprecise dosimeter. The independence assumption implies that knowledge of the value of the *measurement error* alone would not help in any way to predict the size of the *true dose*. In other words, the *measurement error* is uninformative noise. The model is appropriate for other types of dose uncertainties if they add a similar uninformative noise to the dosimetry process. One consequence of the independence assumption, which is sometimes helpful for understanding this model, is that the variance of the *observed doses* in a population of subjects is greater than the variance of their *true doses* (it includes the extra noise; the variance of the *observed doses* equals the variance of the *true doses* plus the variance of the *measurement errors*).

Berkson (*8*) pointed out that errors arising as deviations from an experimentally *controlled* variable follow a different model and have different consequences for regression estimation. Suppose, for example, that a setting on an X-ray machine is used to control doses applied to laboratory animals and that a single organ-specific radiation dose is derived from this setting, but that the actual *true doses* of animals exposed to the same setting differ because of different exposure distances (due to different animal sizes and orientations during the X-ray exposure). All *observed doses* corresponding to one setting will be identical, but the *true doses* will vary about this due to ignored *individual peculiarities*. This model may also be appropriate more generally whenever a single dose is used for a group of subjects, as for example when a uranium miner's *observed dose* is taken to be the average dose for all miners in a specific work location and period. His *true dose* differs from this because of his *individual peculiarity*—his dose's departure from the average. In this model, in contrast to that for classical measurement error, the variance of *true doses* in a population is greater than the variance of the *observed doses*—because variability due to *individual peculiarities* has been excluded from the latter (the variance of the *true doses* equals the variance of the *observed doses* plus the variance of the *individual peculiarities*).

One difficulty in model assessment is that dose uncertainties in radiation epidemiology are rarely due to the archetypal examples for classical and Berkson errors—an imprecise measuring device and a controlled variable. Some helpful key words are ''noise'', ''estimation error'', and ''sampling error'' for the classical error model and ''predicted value'', ''simplification'', ''ignoring individual peculiarities'', or ''grouping'' for the Berkson model. Classical measurement errors are usually present if the dosimetry system includes measurements for individuals or subgroups that are subject to error, e.g. dosimeter readings, questionnaire responses, and environmental sampling results (and, in the case of the Japanese atomic bomb survivors, individual location and shielding). Berkson errors are usually present when doses for groups of subjects are estimated as average values or predicted values from some model (such as a computer model or a regression model). In many dosimetry systems, both classical and Berkson errors are present.

Dose Reconstruction Models with Uncertain Assumptions and Imprecise Inputs

Reeves *et al.* (*9*) and Cox *et al.* (*10*) presented a model with a Berkson component due to the use of a group mean as an ideal estimate of *true dose* and a classical component due to the replacement of that mean by an uncertain estimate. A model for radon dose in the *i*th house in a neighborhood, for example, is

$$\textit{true dose}_i = \mu + \textit{individual peculiarity}_i,$$

$$\textit{observed dose}_i = \hat{\mu},$$

where μ is a neighborhood mean and $\hat{\mu}$ is an uncertain estimate of the mean. This may be re-expressed as $\textit{observed dose}_i = \textit{true dose}_i + (\hat{\mu} - \mu) - \textit{individual peculiarity}_i$.

In an extension of the model of Reeves *et al.*, μ can be

replaced by a linear regression model for dose as a function of a dose-predictor variable, Z_i:

$$\textit{true dose}_i = Z_i'\alpha + \textit{individual peculiarity}_i,$$

$$\textit{observed dose}_i = Z_i'\hat{\alpha},$$

where α represents a set of regression coefficients. In this case, the two sources of uncertainty—the *individual peculiarity* and the *sampling error*, $(Z_i'\hat{\alpha} - Z_i'\alpha)$—are familiar components of variability used in the construction of regression prediction intervals (*11*). If, as is standard in linear regression, the Z_i's are taken to be fixed at their observed values, then the *individual peculiarities* are independent of *observed doses* (and hence Berkson), and the sampling errors are independent of *true doses* (and hence classical).

These relatively simple settings are intended to provide insight into the more realistic scenario in which the *observed dose* is from a complex dose reconstruction:

$$\textit{true dose}_i = f(Z_i, P) + \textit{individual peculiarity}_i, \quad (5)$$

$$\textit{observed dose}_i = f(\hat{Z}_i, \hat{P}), \quad (6)$$

in which Z_i is a set of dose predictor variables (such as distance of the *i*th atomic bomb survivor from the point of detonation, i.e. *exposure distance*), P is a set of true values of numerical quantities used in the dosimetry (such as radiation yield of the atomic bomb), $f(Z_i,P)$ represents the dosimetry ''formula'' for computing an estimated dose from dose-predictor variables Z_i and parameters P, and $\hat{Z}_i$ and $\hat{P}$ are uncertain estimates or guesses of Z_i and P. The *individual peculiarity* again reflects the inability of the dosimetry to predict a subject's dose exactly, even if the exact values of Z_i and P are available, because not all factors can be modeled.

The assessment of classical and Berkson models becomes more difficult in this case, partly because the effects of the uncertainties in $\hat{Z}_i$ and $\hat{P}$ are not easily separated and partly because the probability models are subjective. Nevertheless, we suspect that the predominant effect of uncertainties in judgmental parameter estimates, $\hat{P}$, is a classical error in the *observed dose*, as would be the case if $\hat{P}$ was an estimate based on real data (reflecting ''noise'' in expert judgment), and that the predominant effect of uncertainty in $\hat{Z}_i$ is a classical or Berkson error in *observed dose* according to whether the uncertainty in Z_i itself is classical or Berkson.

Both classical and Berkson models are realistic possibilities for dose-predictor variables. If the atomic bomb survivor dose-predictor variable, *exposure distance*, for example, is uncertain because of survivor recollection errors, then its effect follows the classical model because the recollection errors add noise. On the other hand, if there is uncertainty in *exposure distance* because researchers assigned a single distance to survivors in a general area, then the uncertainty follows the Berkson model because *individual peculiarities* are ignored.

It should be noted that the single uncertainty $\hat{P} - P$ is a component of the dose *error* for all subjects. Stram and Kopecky (*12*) consequently refer to this as a ''shared error.'' The impact may be the same for all individuals, such as when P is the radiation yield of the atomic bomb, so that all doses are similarly over- or underestimated. The impact of shared errors may differ among individuals, though, such as when P is the exact detonation location of the bomb or the rate of radiation intake per volume of milk consumed by a subject in a nuclear fallout study.

BIASES INDUCED BY MULTIPLICATIVE CLASSICAL AND BERKSON DOSE UNCERTAINTIES IN LINEAR AND LINEAR-QUADRATIC DOSE–RESPONSE MODELS

Historically, research into the statistical problem of regression with uncertain explanatory variables has focused on ordinary linear regression with additive measurement errors. It is well known that the estimated slope in simple linear regression is biased toward zero if the uncertainty follows the classical model and it is unbiased if the uncertainty follows a Berkson model (*13–15*). For nonlinear regression and non-additive error structures, these statements are roughly true, but the degree of roughness may be consequential (*3*).

This section details the principal forms of bias in dose–response estimation due to multiplicative Berkson and classical errors. Some attention is given here to the curvature that may be induced or masked by dose uncertainties. While most radiation epidemiology studies have insufficient power for detecting and quantifying curvature, there are nevertheless two reasons for its examination. First, the extent to which an observed dose–response *line* has a slope that is larger or smaller than a true dose–response line depends on a combined effect of an induced curvature and the degree of skewness in the population of *true doses*, as will be shown. Second, the practice of adjusting estimated cancer risks by a dose and dose-rate effectiveness factor (DDREF) is motivated by a belief in upward dose–response curvature, as observed in animal studies. A natural question is whether such upward curvature, in the A-bomb survivor data for example, could be present but masked by Berkson or classical uncertainties.

Biases in Linear Dose–Response Models Due to Single Sources of Multiplicative Uncertainty

Of particular interest in radiation epidemiology are binary regression models and hazard regression models (i.e. dose–response models for age-specific incidence or mortality rate models estimated from subject survival times) that are linear or linear-quadratic functions of dose. Consider first a binary response variable representing cancer incidence or mortality, and suppose that the probability of this response is a linear function of *true dose*,

$$Pr(\textit{response} \mid \textit{true dose}) = \alpha + \beta \times \textit{true dose}. \quad (7)$$

Then the observed dose–response model, i.e. the regression of the binary response on the *observed dose*, must be

$$Pr(\textit{response} \mid \textit{observed dose}) = \alpha + \beta\, Ave(\textit{true dose} \mid \textit{observed dose}), \quad (8)$$

where *Ave*(*true dose*|*observed dose*) is the conditional mean of true doses for the subpopulation of subjects with a given value of *observed dose*. This follows from elementary properties of conditional means; see also ref. (*4*). Using more sophisticated arguments, a similar statement can be made about hazard models: If $\alpha + \beta \times$ *true dose* is a model for the hazard rate as a function of *true dose*, and if the disease outcome of interest is fairly rare, then the hazard rate as a function of the *observed dose* is approximately $\alpha + \beta$ *Ave*(*true dose*|*observed dose*) (*16–18*).

Because of these results, the bias induced by the dose uncertainties can be investigated by examining the extent to which *Ave*(*true dose*|*observed dose*) differs from *observed dose*. If *Ave*(*true dose*|*observed dose*) = *observed dose*, then the coefficient of *observed dose* in Eq. (8) is the same as the coefficient of *true dose* in the model of interest, Eq. (7), so an essentially unbiased estimation procedure using the *observed doses* will yield essentially unbiased estimates of the coefficient of interest. For the uncertainty models of interest here, however, *Ave*(*true dose*|*observed dose*) $\neq$ *observed dose*. (It is important to note that this inequality is not an indictment of the dosimetry process, but rather a feature of conditional means in bivariate distributions, related to the notion of ''regression toward the mean'', which is present even when the dosimetry is accurate.)

Multiplicative Berkson uncertainties may inflate the dose–response slope. Consider the following multiplicative Berkson model:

$$\textit{true dose} = \textit{observed dose} \times \textit{peculiarity factor}, \quad (9)$$

or equivalently

$$\log(\textit{true dose}) = \log(\textit{observed dose}) + \log(\textit{peculiarity factor}), \quad (10)$$

with the assumptions that the logs of the individual *peculiarity factors* follow a normal distribution with mean 0 and variance σ_B^2 (''*B*'' for Berkson), and are independent of *observed dose*. The normality assumption is used for convenience here to demonstrate the effect in a setting where it can be derived with mathematical statistics. In fact,

$$Ave(\textit{true dose} \mid \textit{observed dose}) = e^{\sigma_B^2/2} \times \textit{observed dose}, \quad (11)$$

so if the slope in the ''true'' dose–response model is β, the slope in the observed dose–response model is $e^{\sigma_B^2/2}\,\beta$ [following from Eq. (8), for example]. The bias induced by the Berkson errors in the estimated coefficient of *true dose* is therefore $(e^{\sigma_B^2/2} - 1)\beta$. A belief that true and observed doses typically differ by about 30% (a coefficient of variation of 0.3, which translates to $\sigma_B = 0.29$), for example, implies a 4% overestimation of the coefficient of dose. This bias is a consequence of the assumption that the mean of the logarithms of the *peculiarity factors* in Eq. (10) is zero, which implies that the mean of the peculiarity factors in Eq. (9) is greater than 1. (The median of the peculiarity factors *is* 1. The mean is greater than the median in a positively skewed distribution.) It is debatable whether the correct specification is that the mean of this *peculiarity factor* is 1 or whether the mean of the logarithm of the *peculiarity factor* is 0, but we believe the latter is usually more natural. (For a heuristic argument, suppose the observed doses are either equal to the true doses, off by a factor of 2, or off by a factor of 3, so that the *peculiarity factors* can be 1/3, 1/2, 1, 2 and 3. Notice that the median of these numbers is 1 but that the mean is larger. The symmetry is reflected on the log scale, for which the mean and median are both zero.) In any case, this bias—if present—may be handled rather simply by redefining *observed doses* as the right hand side of Eq. (11), as discussed in a later section.

Multiplicative classical uncertainties usually induce downward curvature. Suppose now that

$$\log(\textit{observed dose}) = \log(\textit{true dose}) + \textit{measurement error},$$

where *measurement error* is taken to be independent of *true dose*, and normally distributed with mean 0 and variance σ_C^2 (''*C*'' for classical). Whereas the Berkson error generates a probability distribution of *true doses* for a given *observed dose*, thus making the computation of *Ave*(*true dose*|*observed dose*) straightforward, the classical error generates a probability distribution for *observed doses* for a given *true dose*. Because of this, consideration of *Ave*(*true dose*|*observed dose*) for assessing bias using Eqs. (7) and (8) is possible only if additional assumptions are made about the distribution of *true doses* in the population of study subjects. If, for example, the *true doses* are lognormally distributed in a population of subjects, so that log(*true dose*) is normal with variance σ^2, then

$$Ave(\textit{true dose} \mid \textit{observed dose}) = \{Ave(\textit{true dose})\}^{1-R^2}(\textit{observed dose})^{R^2}, \quad (12)$$

where $R^2 = \sigma^2/(\sigma^2 + \sigma_C^2)$ (which is the squared correlation of log *true dose* and log *observed dose*, and is between 0 and 1) and *Ave*(*true dose*) is the mean of the *true doses* in the study population. By inserting Eq. (12) in Eq. (8), it is evident that while the true dose–response relationship may be linear, the apparent relationship based on observed doses will be curved. The second derivative of the expression for *Ave*(*true dose*|*observed dose*) in Eq. (12) with respect to *observed dose* is negative, indicating that downward curvature is induced, as has also been shown in refs. (*9*) and (*19*).

Figure 1 provides an illustration of the effect of multiplicative classical uncertainties on simulated data. *True*

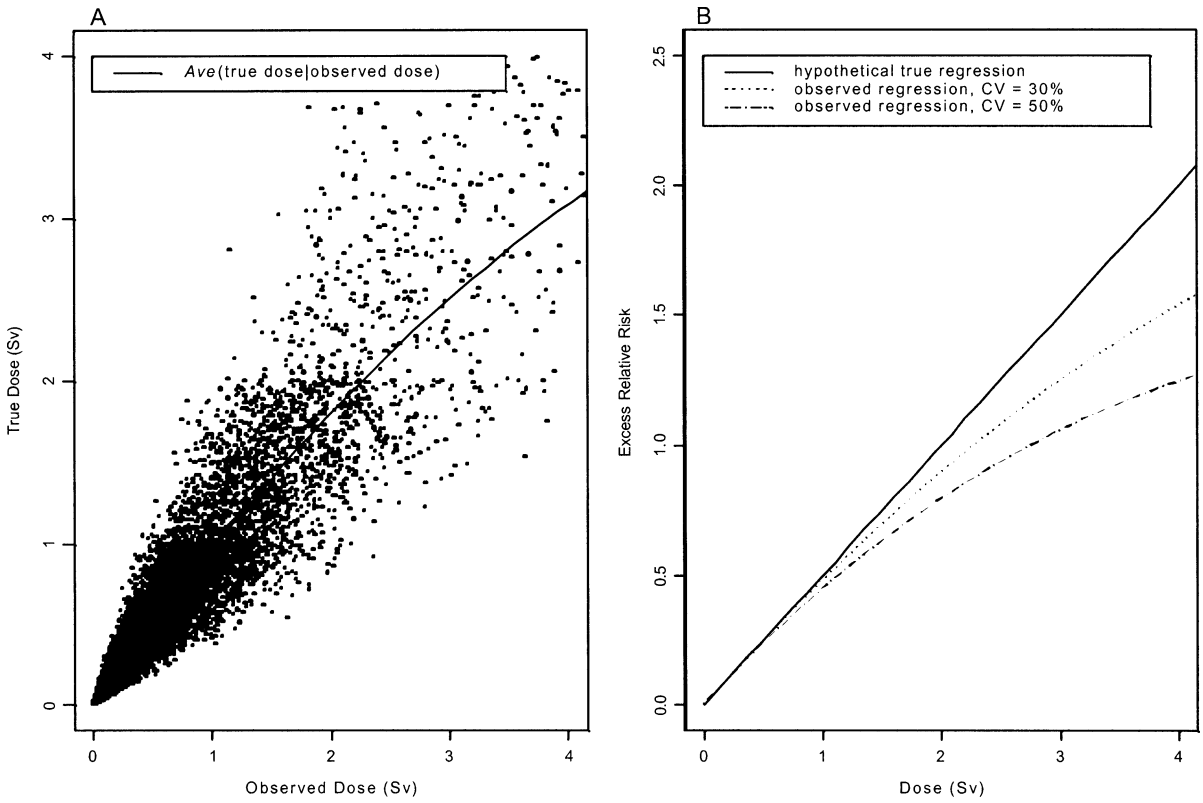

FIG. 1. Illustration of biases induced by dose uncertainties that follow the multiplicative classical model. Panel A shows 86,000 hypothetical *true doses* chosen to roughly match the reported doses of the Japanese A-bomb survivors, and *observed doses* simulated by multiplying these by lognormally distributed error factors. The curve for *Ave*(*true dose*|*observed dose*) is the result of a polynomial fit to the regression of *true* on *observed dose*. Panel B shows a hypothetical excess relative risk model, Excess Relative Risk = 0.5 × *true dose* (solid line), and two associated observed excess relative risk models, = 0.5 × *Ave*(*true dose*|*observed dose*), corresponding to measurement error coefficients of variation of 30% and 50%.

doses were simulated from a skewed distribution that roughly matches the doses for the Japanese A-bomb survivors, based on Table 2 in ref. (*20*). *Observed doses* were generated as the products of these simulated *true doses* and computer-generated, lognormally distributed *measurement error factors*. The plot in panel A shows that *Ave*(*true dose*|*observed dose*) is curved. The effect on the shape of the dose–response curve is illustrated in panel B, demonstrating the potential for incorrectly concluding that there will be downward dose–response curvature if multiplicative classical measurement errors are ignored.

A question then is what bias results from the estimation of a dose–response *line* if multiplicative classical measurement errors are ignored. Notice in panel B of Fig. 1 that a line that connects the zero point of the graph to the dashed or dotted curves will have a slope less than that of the true dose–response line over the region of large doses and a slope slightly greater than that of the true dose–response line over the region of small doses (less than 0.75 Sv, roughly). An estimate of a dose–response line estimates something like a weighted average of these individual lines at all observed doses. Since a large proportion of the doses are ''small'', it is possible, depending on the degree of skewness of *true doses*, that the estimated line might actually have a positive rather than a negative bias. [More formally, at an observed dose equal to some value D, the slope of the line tangent to the observed dose–response curve is $R^2\{Ave(true\ dose)\}^{1-R^2}D^{R^2-1}\beta$. When D is *Ave*(*true dose*), this slope is $R^2\beta$, which is less than β. As D gets smaller, though, the slope gets larger and may be greater than β.]

Can Multiplicative Uncertainties Mask or Distort Real Dose–Response Curvature?

A separate question is whether Berkson and classical errors may mask or distort some underlying curvature that is indeed present in the true dose–response relationship. Suppose that the true regression (binary or hazard) is $\alpha + \beta \times true\ dose + \gamma \times true\ dose^2$. Then the observed regression is

$$\alpha + \beta \times Ave(true\ dose \mid observed\ dose) + \gamma \times Ave(true\ dose^2 \mid observed\ dose),$$

where *Ave*(*true dose*^{2}|*observed dose*) is the mean value of *true dose*^{2} for a subset of subjects with the same value of *observed dose*.

Multiplicative Berkson uncertainties increase apparent curvature. For the lognormal Berkson model in Eq. (10), $Ave(true\ dose^2 \mid observed\ dose) = e^{2\sigma_B^2} \times observed\ dose^2$. Thus the bias in the estimated coefficient of the quadratic

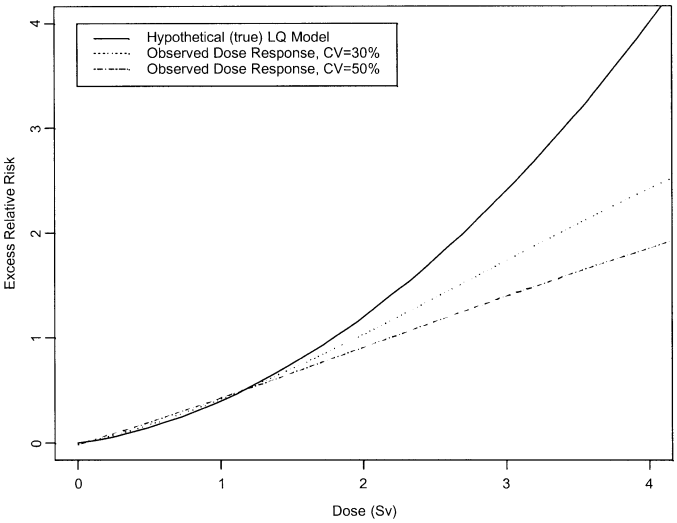

FIG. 2. Illustration of curvature masked by dose uncertainties that follow the multiplicative classical model. The plot shows a hypothetical linear-quadratic (LQ) excess relative risk model, Excess Relative Risk = $0.2 \times$ *true dose* $+ 0.2 \times$ *true dose*^{2}, and two observed excess relative risk models, $= 0.2 \times$ *Ave*(*true dose*$|$*observed dose*) $+ 0.2 \times$ *Ave*(*true dose*$^2|$*observed dose*) induced by lognormal measurement error factors with coefficients of variation of 30% and 50%, on simulated data. The values of *Ave*(*true dose*$|$*observed dose*) were determined as in Fig. 1; the values of *Ave*(*true dose*$^2|$*observed dose*) were similarly obtained by polynomial regression.

term will be $(e^{2\sigma_B^2} \times -1)\gamma$ if the Berkson uncertainties are ignored. Since the term in parentheses is positive, the coefficient of the quadratic term in the observed regression is greater than the one in the true regression, so ignoring multiplicative Berkson uncertainties may cause one to conclude that there is a greater degree of curvature than in fact really exists. A coefficient of variation in the Berkson errors of 30%, for example, corresponds to a 19% overestimation of the magnitude of the quadratic coefficient.

Multiplicative classical uncertainties reduce apparent curvature. A simulated example is used to demonstrate the effect of multiplicative classical uncertainties on the curvature in a true linear-quadratic dose response, using the same simulated *true* and *observed doses* of Fig. 1, but assuming an underlying linear-quadratic relative risk model (Excess relative risk $= 0.2 \times$ *true dose* $+ 0.2 \times$ *true dose*^{2}). Shown in Fig. 2 is this hypothetical true relative risk model and the observed relative risk models induced by lognormal *measurement error factors* with coefficients of variation of 30% and 50%. If there is underlying dose–response curvature like that demonstrated by the solid line, then ignoring multiplicative classical measurement errors may lead one to conclude that the dose response is more like the dotted or dashed line. In this example, the true curvature is masked if multiplicative classical measurement errors are not accounted for. Some formal results are provided in Appendix B for the case that the *true doses* are lognormally distributed, leading to similar conclusions.

Shared Uncertainties from Imprecise Model Terms

We now consider the model of Eqs. (5) and (6) for the specific case that the *observed doses* come from a simple linear regression prediction on the log scale,

$$\log(\textit{true dose}_i) = \alpha_0 + \alpha_1 Z_i + \textit{model error}_i,$$

$$\log(\textit{observed dose}^i) = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i,$$

where Z_i is a dose-predictor variable for subject i; the α's are regression coefficients; and the $\hat{\alpha}$'s are their estimators, derived independently from a ''calibration'' data set. Such a model may be appropriate for organ-specific doses to patients clinically exposed to X rays, with Z representing X-ray machine settings and with the $\hat{\alpha}$'s being estimates from studies of ''phantoms'' subjected to similar X rays [see, for example, ref. (*18*)], but the intent here is primarily to examine this model for understanding analogous effects in more complex dose reconstructions. As discussed previously, the *model error* follows the Berkson model and the component of dose uncertainty due to estimation of the $\hat{\alpha}$'s follows the classical model. These classical *measurement errors* are ''shared'' in the sense that the same term $\hat{\alpha} - \alpha$ appears in the error term for all individuals, although with different impacts depending on the value of Z_i.

Appendix C contains some theoretical results that show a combined effect of the two uncertainties to be similar to what would be expected from the results for Berkson and classical errors individually. In particular, if there is no uncertainty in the $\hat{\alpha}$'s, then R^2 is 1 and the bias reduces to Eq. (11) with a Berkson effect alone. The effect of uncertainty due to imprecision in the estimated prediction equation has much the same form as Eq. (12), inducing downward curvature.

For illustration, computer simulations were performed with $\hat{\alpha}$'s estimated from an independent calibration data set consisting of m pairs of Z's and *true doses* for several choices of m. Figure 3 shows the percentage bias in an estimated slope of a binary regression model estimated from a primary data set of 5000 binary responses and Z's. The *observed doses* were computed as $\hat{\alpha}_0 + \hat{\alpha}_1 Z$, using the estimated $\hat{\alpha}$'s from the independent calibration data set. Further details about the simulation conditions are provided in Appendix D.

For large enough m, only the Berkson bias is present so that the estimated slope will have expectation $e^{\sigma_B^2/2}$, which is 1.13β for the conditions here (hence the horizontal line at 13% in Fig. 3). For smaller m's, there is an additional positive bias, which can be substantial. The reason that the bias is positive, in contrast to the attenuation typically expected from classical measurement errors, is the preponderance of subjects with lower than average doses, as discussed earlier. Broad conclusions cannot be drawn from this simulation, but the substantial biases and the prevalence of real data problems with analogous structure suggest a need for further study, including an understanding of the effect of the correlated measurement errors on estimation tech-

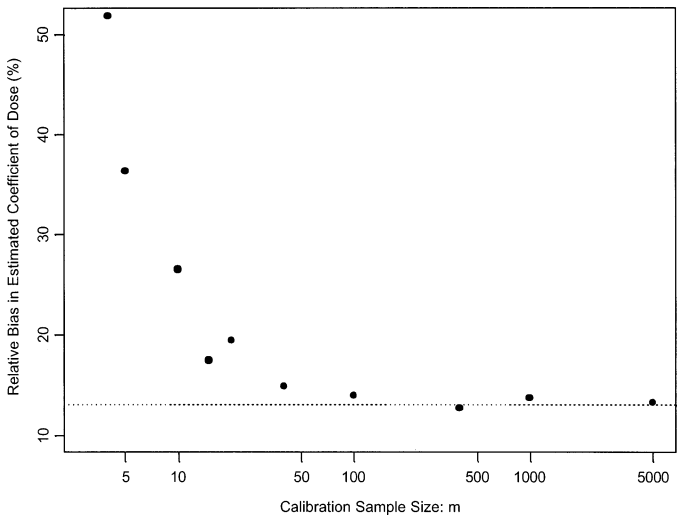

FIG. 3. Simulation results showing the bias in the estimated slope of a linear dose–response model for a binary response, when *observed doses* are predicted values from a regression of *true dose* on a dose-predictor variable *Z*, with estimated coefficients of the prediction model estimated from an independent calibration data set of size *m*. Shown is the mean bias, over a large number of simulations, as a percentage of the true slope. The horizontal dashed line is at 13%, the expected bias due to the Berkson uncertainties alone, from Eq. (A2), Appendix C.

niques and an understanding of the effect of the shared classical error on the estimation of linear-quadratic models. An investigation of shared uncertainties along different lines and for other consequences was given in ref. (*12*).

SOME NOTES ON STATISTICAL METHODS FOR INFERENCE IN THE PRESENCE OF DOSE UNCERTAINTY

There are several approaches for estimation that account for dose uncertainties. The method of maximum likelihood (or maximum partial likelihood for censored survival times) is theoretically optimal in several ways, but it requires that all probability distributions involved be specified completely and, for the types of problems considered here, requires some nontrivial computer programming. General references are (*2*), (*3*) and (*6*). An attractive alternative is to apply essentially the method that would have been appropriate if doses were known precisely but with the missing *true doses* replaced by *Ave*(*true dose*|*observed dose*). This method, which has its roots in the work by Armstrong (*21*) for generalized linear models and Prentice (*16*) for proportional hazard regression, was popularized for logistic regression by Rosner, Willet and Spiegelman (*22*) and was extended and studied as the regression calibration method (*3*).

Radiation epidemiology studies that have explicitly dealt with radiation uncertainties have generally used some version of one of these two methods. The regression calibration method has been used to take account of classical errors in analyses of A-bomb survivors (*4*), persons exposed to radon in houses in Sweden (*23*) and China (*24*), and Colorado Plateau uranium miners exposed to radon (*25*). Errors in estimates of exposure to radon in houses have also been investigated in a study in Southwest England (*9, 26*) using an approximate likelihood analysis, with characteristics similar to regression calibration. In addition to accounting for classical error, this study accounted for the Berkson error resulting from the use of imputed values for addresses without radon measurements. Bayesian approaches (which extract information from the data in the same way as maximum likelihood) were used by Thomas (*27*) in a case–control study of leukemia in Utah residents exposed to radioactive fallout from atmospheric nuclear tests and also in a cohort study of thyroid disease in persons exposed to fallout as children and by Mallick *et al.* (*28*) for studying thyroid disease in persons exposed to fallout from the Nevada test site. Schafer *et al.* (*18*) investigated both maximum likelihood and regression calibration in a reanalysis of thyroid cancer risk in Israeli tinea capitis patients treated with X rays, finding little difference between the two.

Accounting for dose measurement error increased the estimated risk coefficients in most of the studies noted above (about 10% for A-bomb survivors, 50–100% in the residential radon studies, 60% for the Colorado miners, 30% for the Utah fallout leukemia study, and 100% for the Utah fallout thyroid study). The relative uncertainty was also increased in studies that used various approaches to address this (*9, 24, 27, 28*). In the Colorado miner study, taking account of error decreased both the magnitude and the statistical significance of the dose-rate effect. In this study, errors in exposures received at high dose rates were larger (on an additive scale) than those received at low dose rates. In the Hanford Thyroid Study, taking account of error reduced the estimated power in a specified situation of interest from 0.80 to 0.71 (*12*). Adjustment for measurement error in the tinea capitis study had only a negligible effect on dose–response estimation or on inferences on the modifying effects of age at first exposure, time since exposure and other factors (*18, 29*).

Because of the results showing that the observed linear dose–response models for binary responses and hazard functions have the same form as the true dose–response models with *true dose* replaced by *Ave*(*true dose*|*observed dose*), it follows that maximum likelihood and regression calibration produce equivalent estimates for these special cases. If the dose response follows a linear-quadratic function, then the equivalence still holds if, in using regression calibration, $true\ dose^2$ is replaced by *Ave*($true\ dose^2$|*observed dose*).

The equivalence is true only if the conditional means, *Ave*(*true dose*|*observed dose*), are known exactly, though, which is usually not the case. A previous section highlighted the potential dangers of assuming that *Ave*(*true dose*|*observed dose*) is known when in fact it is an estimate based on an estimated prediction equation. This may be a fine point, since using an estimate of *Ave*(*true*

dose|*observed dose*) is a substantial improvement over ignoring uncertainties, but additional study for further refinement of this method is warranted.

Notice that for multiplicative Berkson uncertainties following the lognormal model, *Ave*(*true dose*|*observed dose*) $= e^{\sigma_B^2/2}$ *observed dose*, so that the regression calibration may be handled rather simply with knowledge of the Berkson error variance. For classical measurement errors, as discussed earlier, *Ave*(*true dose*|*observed dose*) is a more complex quantity that depends crucially on the distribution of *true doses* in the study population. See ref. (*30*) for a fuller discussion of this issue and an approach for estimating *Ave*(*true dose*|*observed dose*) that takes this distribution into account but without a need to specify a particular form for it.

There are two modeling extremes in applying regression calibration when *observed dose* is the result of a complex dose reconstruction process. One is to attempt to model the dose reconstruction entirely, accounting for each piece of the dose reconstruction and the uncertainties that arise from each piece, and the other is to simply lump all uncertainties into either a classical or Berkson category and speculate on their overall variances. Examples of the former are in refs. (*12*) and (*31*) and of the latter in ref. (*4*). Sensitivity analyses are important in both approaches.

A useful and popular tool for detailed modeling of dose reconstruction is computer simulation, in which a simulated distribution of possible doses for an individual emerges through simulated uncertainties at various stages of the dose reconstruction process (*31*), permitting the estimation of *Ave*(*true dose*|*observed dose*) for use in the regression calibration approach. This has the advantage that effects of shared uncertainties are handled automatically. There is a potential danger in mistaking classical and Berkson errors, perhaps more so than in other analyses, because of the ease in generating pseudo measurement errors according to one model and the potential confusion between the population distribution of *true dose* given *observed dose* and a hypothetical distribution of *true dose* given *observed dose* for a particular individual. Because of the overall attractiveness of the simulation approach along with the potential pitfalls in keeping various probability distributions straight, further study and clarification of the use of simulated distributions of possible *true doses* for regression calibration and for likelihood or Bayesian analysis [as suggested in ref. (*32*)] are warranted. A recent discussion of this approach is in ref. (*33*).

DISCUSSION

This paper emphasizes the multiplicative versions of the classical and Berkson dose uncertainty models. The classical/Berkson modeling discussion has been explained many times previously in the context of radiation epidemiology [for example, in ref. (*34*)]. Further conceptual clarification is provided here for thinking about the distinction in practice by noting that the formal independence assumptions imply that the classical model is appropriate when the uncertainty is due to uninformative noise and that the Berkson model is appropriate when the uncertainty is due to the ignoring of *individual peculiarities*.

The primary effect of multiplicative Berkson uncertainties on linear and linear-quadratic dose–response models is a possible exaggeration of dose–response curvature. The primary effects of multiplicative classical uncertainties are the introduction of downward curvature and the masking of upward curvature if it is present in the dose–response relationship. The effect of multiplicative classical uncertainties on linear dose–response estimation is difficult to generalize as it depends on the skewness of *true doses* in the study population, in addition to the induced curvature.

The use of *Ave*(*true dose*|*observed dose*) as a substitute for the unavailable *true dose* has been an important part of recent statistical analyses in radiation epidemiology, offering substantial improvement over naïve analyses that ignore dose uncertainties. Given that the available *Ave*(*true dose*|*observed dose*) is only an estimate (or based largely on imprecise expert judgment), though, further improvement may be possible by accounting for the effect of the classical errors involved in estimation or speculation of common model terms and the effects of these shared uncertainties on the statistical inferential tools used. Related to this are similar issues involved when *Ave*(*true dose*|*observed dose*) is determined by simulation.

APPENDIX A

Definition of Statistical Independence

Let X and W represent two random variables, such as the *true dose* and the *measurement error* (in the *observed dose*) for a randomly selected subject in a population of exposed subjects in an epidemiological study. Use of the term ''random variables'' implies that numerical values for X and W would result from random selection. We may consider their possible and likely values through a probability distribution. The marginal probability density for W, $f(w)$, describes the probability of values that W takes in the population, without regard to values of X. The conditional probability density, $f(w|X = x)$, refers to the probability distribution of W for all subjects whose value of X is x. W and X are independent if $f(w|X = x) = f(w)$. In other words, W and X are independent if knowledge of the numerical value for X provides no information for refining knowledge about the possible and likely values of W. One consequence of independence is that the conditional mean of W given X is equal to the marginal mean: $Ave(W|X = x) = Ave(W)$.

APPENDIX B

Effect of Multiplicative Classical Uncertainties on Estimation of a Linear-Quadratic Dose–Response Model

In the classical model with lognormal *measurement error factors* and lognormal *true doses*, $Ave(\text{true dose}^2|\text{observed dose}) = \{Ave(\text{true dose}^2)\}^{1-R^2}(\text{observed dose}^2)^{R^2}$. It can be inferred from this and Eq. (12) that

$$Ave(y \mid \text{observed dose}) = \alpha + \beta\mu^{1-R^2}(\text{observed dose})^{R^2} + \gamma\mu_2^{1-R^2}(\text{observed dose}^2)^{R^2}, \quad \text{(A1)}$$

where μ and μ_2 are the mean and second moment [i.e. *Ave*(*true dose*^{2})] of *true dose* in the population of subjects. A quadratic approximation to Eq. (A1) (obtained by expanding the second term about *observed dose* $= \mu$ and expanding the third term about *observed dose*$^2 = \mu_2$) indicates both a multiplicative and an additive (negative) effect on the coefficient of the squared term. If β and γ are both positive then the coefficient of the squared term in the observed regression will be reduced from γ both by a factor less than 1 and by an additional subtracted amount. Thus the apparent curvature in the observed regression will be less than the true (upward increasing) curvature and may even be curved in the opposite direction.

APPENDIX C

Bias Due to Dose Estimates Arising from an Estimated Regression Prediction Equation

If, in the regression prediction model, the *Z*'s, the *model errors* and the $\hat{\alpha}$'s are all normally distributed, then it can be shown that

$$Ave(\textit{true dose} \mid \textit{observed dose}) = e^{R^2\sigma_B^2/2}\{Ave(\textit{true dose})\}^{1-R^2}(\textit{observed dose})^{R^2}, \quad \text{(A2)}$$

where σ_B^2 is the variance of the (Berkson) *model errors*, and

$$R^2 = \frac{\alpha_1\sigma_Z^2}{[\alpha_1 + \text{var}(\hat{\alpha}_1)]\sigma_Z^2 + \text{var}[\hat{\alpha}_0 + \hat{\alpha}_1 Ave(Z)]}.$$

APPENDIX D

Details of Simulation

In each run of the simulation displayed in Fig. 3, a primary data set was generated as follows: 5000 *Z*'s were generated from a normal distribution with mean 0 and standard deviation 1, 5000 log *true dose*'s were generated from a normal distribution with mean $-1.6 + 0.75Z$ and standard deviation 0.5, and 5000 binary responses were generated as Bernoulli with probability $0.1 + 0.05$ *true dose*. A calibration data set was also simulated: *m Z*'s were generated from the same normal distribution as in the primary data set and *m true doses* were generated from the same normal distribution as in the primary data set. With these sets of simulated variables, the following steps were applied: The regression of log *true dose* on *Z* was estimated by least squares from the calibration data set, the resulting estimated coefficients were applied to the *Z*'s in the primary data set to arrive at predicted doses (which serve as the *observed doses*), and the dose–response regression coefficients were estimated by maximum likelihood estimation of the binary responses on these *observed doses*. These simulation runs were repeated 10,000 times for $m = 4$ and 5; 1000 times for $m = 10$, and 500 times for $m = 15$, 20, 40, 100, 400, 1000 and 5000. For each of these, the mean bias was computed as the mean of the Monte Carlo distribution of the slope estimates in the binary dose–response regression minus the true value (0.05).

REFERENCES

1. B. G. Armstrong, The effects of measurement errors on relative risk regressions. *Am. J. Epidemiol.* **132**, 176–184 (1990).
2. D. G. Clayton, Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In *Statistical Models for Longitudinal Studies of Health* (J. H. Dwyer, P. Lippert, M. Feinleib and H. Hoffmeister, Eds.), pp. 301–333. Oxford University Press, New York, 1990.
3. R. J. Carroll, D. Rupert and L. A. Stefanski, *Measurement Error in Nonlinear Models.* Chapman Hall, London, 1995.
4. D. A. Pierce, D. O. Stram and M. Vaeth, Allowing for random errors in radiation dose estimates for the atomic bomb survivors. *Radiat. Res.* **123**, 275–284 (1990).
5. D. A. Pierce, D. O. Stram, M. Vaeth and D. W. Schafer, Some insights into the errors in variables problem provided by consideration of radiation dose–response analyses for the A-bomb survivors. *J. Am. Stat. Assoc.* **87**, 351–359 (1992).
6. D. C. Thomas, D. Stram and J. Dwyer, Exposure measurement error: Influence on exposure-disease relationships and methods of correction. *Annu. Rev. Publ. Health* **14**, 69–93 (1993).
7. J. H. Lubin, J. M. Samet and C. Weinberg, Design issues in epidemiologic studies of indoor exposure to radon and risk of lung cancer. *Health Phys.* **59**, 807–817 (1990).
8. J. Berkson, Are there two regressions? *J. Am. Stat. Assoc.* **45**, 164–180 (1950).
9. G. K. Reeves, D. R. Cox, S. C. Darby and E. Whitley, Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Stat. Med.* **17**, 2147–2177 (1998).
10. D. R. Cox, S. C. Darby, G. K. Reeves and E. Whitley, The effects of measurement errors with particular reference to a study of exposure to residential radon. In *Uncertainties in Radiation Dosimetry and Their Impact on Dose–Response Analyses* (E. Ron and F. O. Hoffman, Eds.), pp. 139–151. Publication no. 99-4541, National Cancer Institute, Bethesda, MD, 1999.
11. F. L. Ramsey and D. W. Schafer, *The Statistical Sleuth, a Course in Methods of Data Analysis*, 2nd ed. Duxbury Press, Belmont, CA, 2002.
12. D. O. Stram and K. J. Kopecky, Power and uncertainty analysis of epidemiologic studies of radiation-related disease risk in which dose estimates are based on a complex dosimetry system: Some observations. *Radiat. Res.* **160**, 408–417 (2003).
13. A. Madansky, The fitting of straight lines when both variables are subject to error. *J. Am. Stat. Assoc.* **54**, 173–205 (1959).
14. W. G. Cochran, Errors of measurement in statistics. *Technometrics* **10**, 637–665 (1968).
15. A. Stuart, J. K. Ord and S. Arnold, *Kendall's Advanced Theory of Statistics: Classical Inference and the Linear Model (Kendall's Library of Statistics).* Oxford University Press, Oxford, 1999.
16. R. L. Prentice, Covariate measurement error and parameter estimation in a failure time regression model. *Biometrika* **69**, 331–342 (1982).
17. M. S. Pepe, S. G. Self and R. L. Prentice, Further results in covariate measurement errors in cohort studies with time to response data. *Stat. Med.* **8**, 1167–1178 (1989).
18. D. W. Schafer, J. H. Lubin, E. Ron, M. Stovall and R. J. Carroll, Thyroid cancer following scalp irradiation: A reanalysis accounting for uncertainty in dosimetry. *Biometrics* **57**, 689–697 (2001).
19. D. A. Pierce and M. Vaeth, The shape of the cancer mortality dose–response curve for the A-bomb survivors. *Radiat. Res.* **126**, 36–42 (1991).
20. D. L. Preston, Y. Shimizu, D. A. Pierce, A. Suyama and K. Mabuchi, Studies of mortality of atomic bomb survivors, Report 13: Solid cancer and noncancer mortality 1950–1997. *Radiat. Res.* **160**, 381–407 (2003).
21. B. G. Armstrong, Measurement errors in generalized linear models. *Commun. Stat. B* **14**, 529–544 (1985).
22. B. Rosner, W. C. Willet and D. Spiegelman, Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat. Med.* **8**, 1050–1070 (1989).
23. F. Lagarde, G. Pershagen, G. Akerblom, O. Axelson, U. Baverstam, L. Damber, A. Enflo, M. Svartengren and G. A. Swedjemark, Residential radon and lung cancer in Sweden: Risk analyses accounting for random error in the exposure assessment. *Health Phys.* **72**, 269–276 (1997).
24. Z. Y. Wang, J. H. Lubin, L. D. Wang, S. Z. Zhang, J. D. Boice, Jr., H. Z. Cui, S. R. Conrath, Y. Xia and R. A. Kleinerman, Residential radon and lung cancer in a high exposure area in Gansu Province, China. *Am. J. Epidemiol.* **155**, 554–564 (2002).
25. D. O. Stram, B. Langholz, M. Huberman and D. C. Thomas, Correction for exposure measurement error in a reanalysis of lung cancer

mortality for the Colorado Plateau uranium miners cohort. *Health Phys.* **77**, 265–275 (1999).

26. S. Darby, E. Whitley, P. Silcockis, B. Thakrar, M. Green, P. Lomas, J. Miles, G. Reeves, T. Fearn and R. Doll, Risk of lung cancer associated with residential radon exposure in Southwest England: A case-control study. *Br. J. Cancer* **78**, 394–408 (1998).

27. D. C. Thomas, The Utah fallout study: How uncertainty has affected estimates of dose-response. In *Uncertainties in Radiation Dosimetry and Their Impact on Dose–Response Analyses* (E. Ron and F. O. Hoffman, Eds.), pp. 217–224. Publication No. 99-4541, National Cancer Institute, Bethesda, MD, 1999.

28. B. Mallick, F. O. Hoffman and R. J. Carroll, Semiparametric regression modeling with mixtures of Berkson and classical error, with application to fallout from the Nevada test site. *Biometrics* **58**, 13–20 (2002).

29. J. H. Lubin, D. W. Schafer, E. Ron, M. Stovall and R. J. Carroll, A reanalysis of thyroid neoplasms in the Israeli tinea capitis study accounting for dose uncertainties. *Radiat. Res.* **161**, 359–368 (2004).

30. D. A. Pierce and A. M. Kellerer, Adjusting for covariate errors with nonparametric assessment of the true covariate distribution. *Biometrika* **91**, 863–876 (2005).

31. F. O. Hoffman, Environmental dose reconstruction: How large can uncertainty be when models take the place of measurements? In *Uncertainties in Radiation Dosimetry and Their Impact on Dose–Response Analyses* (E. Ron and F. O. Hoffman, Eds.), pp. 5–23. Publication No. 99-4541, National Cancer Institute, Bethesda, MD, 1999.

32. R. J. Carroll, Risk assessment with subjectively derived doses. In *Uncertainties in Radiation Dosimetry and Their Impact on Dose–Response Analyses* (E. Ron and O. Hoffman, Eds.), pp. 37–48. Publication No. 99-4541, National Cancer Institute, Bethesda, MD, 1999.

33. K. J. Kopecky, S. Davis, T. E. Hamilton, M. S. Saporito and L. E. Onstad, Estimation of thyroid radiation doses for the Hanford thyroid disease study: Results and implications for statistical power of the epidemiological analyses. *Health Phys.* **87**, 15–32 (2004).

34. E. Ron and F. O. Hoffman, Eds., *Uncertainties in Radiation Dosimetry and Their Impact on Dose–Response Analyses.* Publication No. 99-4541, National Cancer Institute, Bethesda, MD, 1999.